Blind hyperspectral unmixing using an Extended Linear Mixing Model to address spectral variability

Lucas Drumetz, Miguel-Angel Veganzones, *Member, IEEE*, Simon Henrot, Ronald Phlypo, *Member, IEEE*, Jocelyn Chanussot, *Fellow, IEEE*, and Christian Jutten, *Fellow, IEEE*

This document contains some supplementary material for the paper "Blind hyperspectral unmixing using an Extended Linear Mixing Model to address spectral variability". The contents are organized as follows: Section I contains all the regularization parameters for all the competing algorithms, for all the tested datasets. Section II presents some complementary results on the first synthetic dataset, namely a study on the effect of smoothing the scaling factors and a sensitivity analysis to the regularization parameters. Section III contains all the optimization details for the minimization of the Augmented Lagrangian w.r.t $\mathbf{u}$, $\boldsymbol{\mu}$, $\mathbf{v}$ and $\mathbf{d}$.

I. Regularization parameters for all the competing algorithms

This Section contains the regularization parameters used for each competing algorithm, and each dataset. The parameters for the synthetic data are gathered in Table I, and the parameters for the real datasets are stored in Table II. $\lambda_{\mathcal{L}_1}$ stands for the sparsity regularization parameter in SUnSAL and SUnSAL-TV. For the PLMM, on the syntetic datasets, the parameters α, β and γ are regularization parameters associated to a Tikhonov regularization on the abundances, on a mutual distance penalization on the endmembers and on a penalization of the spectral variability power (see [1] for details). On the real datasets, they are associated to a spatial regularization on the abundances, to a distance of the endmembers to $\mathbf{S}_0$, and to the spectral variability power, respectively.

II. Complementary Results on the first synthetic dataset

A. Interest of smoothing the scaling factors

Here we describe some additional results to show the relevance of smoothing the scaling factors. We show in Fig. 1 a plot of the estimated scaling factors against the true ones, for one of the endmembers of the image, and for two algorithms: ELMM and ELMM-Aψ. This direct comparison between estimated and true scaling factors only makes sense if the reference endmember matrix $\mathbf{S}_0$ is the same in both cases. Hence, here we assumed the reference endmembers were known. The red dots in Fig. 1 show that without any spatial regularization, the ELMM is able here to estimate spectral variability only when the abundance coefficient in one pixel for the considered material is above a certain value (here around 0.3). Otherwise, the estimated values stay close to their initial value 1. We had already mentioned this phenomenon above, and explained it geometrically. However, with the spatial regularization on the scaling factors, we see that not only are we able to reduce the estimation error for the red dots in most cases, we are also in general able to stir the other pixels with lower abundance values towards a more accurate spectral variability estimation, using the spatial information. As expected, the spatial smoothness of the scaling factor then helps estimating spectral variability in more heavily mixed pixels than with a pixelwise approach. Note that here, only the spatial regularization on the scaling factors impacts the shape of this plot, while the TV on the abundances does not play an important role. Notwithstanding, the latter is still very useful on its own for the abundance estimation, and as shown in the quantitative results, it is complementary to the spatial regularization on the scaling factors since a better scaling factor estimation allows in turn a better abundance estimation.

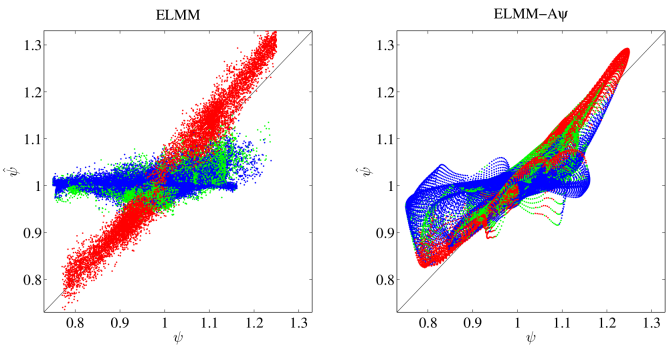

Fig. 1. Plot of the estimated scaling factors against the true ones (assuming $\mathbf{S}_0$ is known), for ELMM and ELMM-Aψ, corresponding to the one of the endmembers of the results (second column of Figs. 4 and 5 of the main paper). Each dot represents a pixel in the image, and its color refers to the value of a_{2i}, where i denotes the pixel index. Red corresponds to $a_{2i} \geq 0.3$, green corresponds to $0.1 < a_{2i} < 0.3$, and blue corresponds to $a_{2i} \leq 0.1$.

L. Drumetz, S. Henrot, and C. Jutten are with the Department Image and Signal (DIS), GIPSA-lab, Université Joseph Fourier, F-38402 Saint Martin d'Hères Cedex, France. Phone (e-mail: lucas.drumetz@gipsa-lab.fr; simon.henrot@gipsa-lab.fr; christian.jutten@gipsa-lab.fr)

M.A. Veganzones is with the Department Image and Signal (DIS), GIPSA-lab, CNRS, F-38402 Saint Martin d'Hères Cedex, France. Phone (e-mail: miguel-angel.veganzones@gipsa-lab.fr

R. Phlypo and J. Chanussot are with the Department Image and Signal (DIS), GIPSA-lab, Grenoble-INP, F-38402 Saint Martin d'Hères Cedex, France(e-mail: ronald.phlypo@gipsa-lab.fr; jocelyn.chanussot@gipsa-lab.fr)

J. Chanussot is also with the Faculty of Electrical and Computer Engineering, University of Iceland, Reykjavik, Iceland.

J. Chanussot and C. Jutten are members of Institut Universitaire de France.

This work was partially supported by the European Research Council under the European Community's Seventh Framework Programme FP7/2007–2013, under Grant Agreements no.320684 (CHESS project) and no.320594 (DECODA project).

	BUNDLES + SUnSAL		BUNDLES + SUnSAL-TV		BUNDLES + FDN-TV		CLSU-TV		PLMM		S-CLSU-TV		ELMM		ELMM-Aψ	
$\lambda_{\mathcal{L}_1}$	2.10^{-3}	5.10^{-4}	2.10^{-3}	10^{-5}	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
λ_{TV}	$\times$	$\times$	2.10^{-3}	5.10^{-4}	3.10^{-3}	10^{-4}	5.10^{-4}	3.10^{-4}	$\times$	$\times$	4.10^{-3}	3.10^{-4}	$\times$	$\times$	$\times$	$\times$
λ_S	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$6.3.10^{-2}$	$\times$	$6.3.10^{-2}$	4
λ_A	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	0	$\times$	5.10^{-3}	$1.5.10^{-2}$
λ_ψ	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	0	$\times$	9.10^{-1}	4.10^{-1}
α	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	10^{-5}	10^{-5}	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
β	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$4.9.10^{-3}$	$4.7.10^{-3}$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
γ	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	1	1	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$

TABLE I
REGULARIZATION PARAMETERS FOR ALL THE ALGORITHMS CONCERNED, FOR THE FIRST SYNTHETIC DATASET (LEFT CELL OF EACH COLUMN) AND THE SECOND SYNTHETIC DATASET (RIGHT CELL OF EACH COLUMN).

	BUNDLES + SUnSAL		PLMM		ELMM-Aψ	
$\lambda_{\mathcal{L}_1}$	5.10^{-4}	5.10^{-4}	$\times$	$\times$	$\times$	$\times$
λ_S	$\times$	$\times$	$\times$	$\times$	0.4	0.4
λ_A	$\times$	$\times$	$\times$	$\times$	5.10^{-3}	5.10^{-3}
λ_ψ	$\times$	$\times$	$\times$	$\times$	10^{-3}	5.10^{-3}
α	$\times$	$\times$	$1.4.10^{-3}$	$3.1.10^{-4}$	$\times$	$\times$
β	$\times$	$\times$	5.10^2	5.10^2	$\times$	$\times$
γ	$\times$	$\times$	1	1	$\times$	$\times$

TABLE II
REGULARIZATION PARAMETERS FOR ALL THE ALGORITHMS CONCERNED, FOR THE HOUSTON DATASET (LEFT CELL OF EACH COLUMN) AND THE CUPRITE DATASET (RIGHT CELL OF EACH COLUMN).

B. Sensitivity analysis

In this section we show some results regarding the sensitivity of the proposed method w.r.t. the three regularization parameters to tune. We have run the ELMM algorithm for $\lambda_S \in [10^{-2}, 10^{-1}]$ with steps of 10^{-1}, $\lambda_A \in [10^{-3}, 10^{-2}]$ with steps of 10^{-2}, $\lambda_\psi \in [1, 10]$ with steps of 1. In order to visualize the sensitivity of the algorithm to the regularization parameters, we have plotted a well chosen isosurface of the three-variable functions given by aRMSE and sRMSE. The chosen values were 0.0215 for aRMSE and 0.044 for sRMSE. Inside the volume delimited by the surface, the metric is lower than the chosen value. This delimits a 3D domain for the regularization parameters inside which performance remains much better than the competing algorithms. The intersection of the two surfaces would delimit a volume inside which good performance is guaranteed both for abundance and spectral variability estimation. We see that logically, λ_A (resp. λ_Ψ) is a critical parameter for a good abundance (resp. spectral variability) estimation, while the value of λ_S is important for both, and is probably the most critical parameter overall. Still, it can be chosen in a relatively large domain for close to optimal performance.

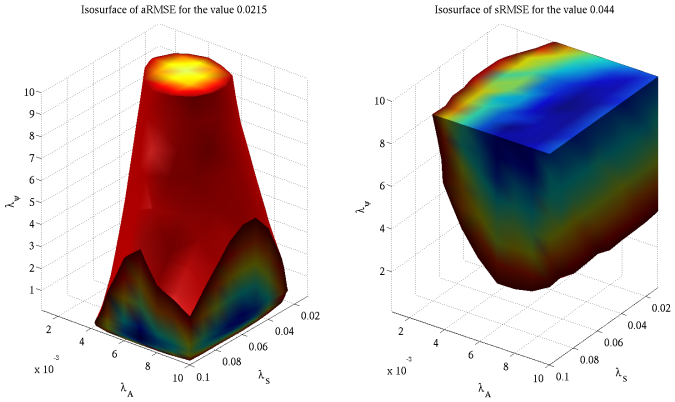

Fig. 2. Isosurfaces for two values of aRMSE = 0.0215 (left) and sRMSE = 0.044 (right) (seen as functions of the three regularization parameters), delimiting a domain inside which the metrics are lower than these two values. For the aRMSE, the color scale ranges from 0.0205 (blue) to 0.0215 (red). For the sRMSE, the color scale goes from 0.0438 (blue) to 0.044 (red).

III. Optimization details for the minimization of the Augmented Lagrangian w.r.t. $\mathbf{u}$, $\boldsymbol{\mu}$, $\mathbf{v}$ and $\mathbf{d}$

We describe below the optimization procedure to minimize the AL of Eq.(1) over each variable. The Augmented Lagrangian writes:

$$\begin{aligned}\mathcal{L}(\mathbf{u},\boldsymbol{\mu},\mathbf{v},\mathbf{d}) &= f(\mathbf{u}) + g(\mathbf{v}) + \frac{\rho}{2}(||\boldsymbol{\Gamma}\mathbf{u} + \boldsymbol{\Lambda}\mathbf{v} - \mathbf{d}||_2^2 - ||\mathbf{d}||_2^2) \\ &= \frac{1}{2}||\mathbf{x} - \mathbf{b}_1||_2^2 + \lambda_A(||\text{vec}^{-1}(\mathbf{b}_2)||_{2,1} + ||\text{vec}^{-1}(\mathbf{b}_3)||_{2,1}) \\ &\quad + \mathcal{I}_{\mathbb{R}_+^{PN}}(\mathbf{b}_4) + \boldsymbol{\mu}^\top(\mathbf{K}\mathbf{u} - \mathbb{1}_N) + \frac{\rho}{2}||\boldsymbol{\Sigma}\mathbf{u} - \mathbf{b}_1 - \mathbf{d}_1||_2^2 \\ &\quad + \frac{\rho}{2}||\mathbf{u} - \mathbf{b}_2 - \mathbf{d}_2||_2^2 + \frac{\rho}{2}||\mathbf{H}_h\mathbf{b}_2 - \mathbf{b}_3 - \mathbf{d}_3||_2^2 \\ &\quad + \frac{\rho}{2}||\mathbf{H}_v\mathbf{b}_2 - \mathbf{b}_4 - \mathbf{d}_4||_2^2 + \frac{\rho}{2}||\mathbf{u} - \mathbf{b}_5 - \mathbf{d}_5||_2^2 - \frac{\rho}{2}||\mathbf{d}||_2^2. \end{aligned} \quad (1)$$

1) Optimization w.r.t $\mathbf{u}$ and $\boldsymbol{\mu}$: This suproblem writes:

$$\begin{aligned}\arg\min_{\mathbf{u},\boldsymbol{\mu}} \sum_{k=1}^{N} \Big(&\frac{\rho}{2}||\mathbf{S}_k\mathbf{u}_k - \mathbf{b}_{1k} - \mathbf{d}_{1k}||_2^2 \\ &+ \frac{\rho}{2}||\mathbf{u}_k - \mathbf{b}_{5k} - \mathbf{d}_{5k}||_2^2 + \frac{\rho}{2}||\mathbf{u}_k - \mathbf{b}_{2k} - \mathbf{d}_{2k}||_2^2 \\ &+ \mu_k(\mathbf{u}_k^\top \mathbb{1}_P - 1)\Big) \end{aligned} \quad (2)$$

and is separable over each pixel. By nulling the gradients of the k^{th} term of Eq. (2) w.r.t. to $\mathbf{u}_k$ and μ_k, we get the following system to solve:

$$\begin{bmatrix} \boldsymbol{\Omega}_k & \mathbb{1}_P \\ \mathbb{1}_P^\top & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u}_k \\ \mu_k \end{bmatrix} = \begin{bmatrix} \boldsymbol{\delta}_k \\ 1 \end{bmatrix}, \quad (3)$$

where

$$\boldsymbol{\Omega}_k = \rho(\mathbf{S}_k^\top\mathbf{S}_k + 2\mathbf{I}_P) \quad (4)$$

and

$$\boldsymbol{\delta}_k = \rho\mathbf{S}_k^\top(\mathbf{b}_{1k} + \mathbf{d}_{1k}) + \rho(\mathbf{b}_{2k} + \mathbf{d}_{2k} + \mathbf{b}_{5k} + \mathbf{d}_{5k}). \quad (5)$$

Finally, by introducing the scalar quantity $s_k = \mathbb{1}_P^\top\boldsymbol{\Omega}_k^{-1}\mathbb{1}_P$ (which is simply the sum of all entries in Ω_k^{-1}), and using the

block matrix inversion formula, we get the update rule for $\mathbf{u}_k$ and μ_k:

$$\begin{bmatrix}\mathbf{u}_k\\ \mu_k\end{bmatrix} \leftarrow \frac{1}{s_k}\begin{bmatrix}\mathbf{\Omega}_k^{-1}(s_k\mathbf{I}_P - \mathbb{1}_P\mathbb{1}_P^\top\mathbf{\Omega}_k^{-1}) & \mathbf{\Omega}_k^{-1}\mathbb{1}_P\\ \mathbb{1}_P^\top\mathbf{\Omega}_k^{-1} & -1\end{bmatrix}\begin{bmatrix}\boldsymbol{\delta}_k\\ 1\end{bmatrix}. \tag{6}$$

2) Optimization w.r.t. $\mathbf{b}_1$*:* Here, we need to solve the following problem :

$$\arg\min_{\mathbf{b}_1} \quad \sum_{k=1}^{N}\frac{1}{2}||\mathbf{x}_k - \mathbf{b}_{1k}||_2^2 + \frac{\rho}{2}||\mathbf{S}_k\mathbf{u}_k - \mathbf{b}_{1k} - \mathbf{d}_{1k}||_2^2. \tag{7}$$

Once again, this is separable over all pixels, and the update rule for $\mathbf{b}_{1k}$ is:

$$\mathbf{b}_{1k} \leftarrow \frac{1}{1+\rho}(\mathbf{x}_k - \rho\mathbf{d}_{1k} + \rho\mathbf{S}_k\mathbf{a}_k). \tag{8}$$

3) Optimization w.r.t. $\mathbf{b}_2$*:* The problem to solve for $\mathbf{b}_2$ is:

$$\begin{aligned}\arg\min_{\mathbf{b}_2} \quad & \frac{\rho}{2}||\mathbf{u} - \mathbf{b}_2 - \mathbf{d}_2||_2^2\\ & + \frac{\rho}{2}||\mathbf{H}_h\mathbf{b}_2 - \mathbf{b}_3 - \mathbf{d}_3||_2^2 + \frac{\rho}{2}||\mathbf{H}_v\mathbf{b}_2 - \mathbf{b}_4 - \mathbf{d}_4||_2^2,\end{aligned} \tag{9}$$

which is readily solved by

$$\begin{aligned}\mathbf{b}_2 \leftarrow & (\mathbf{I}_{PN} + \mathbf{H}_h^\top\mathbf{H}_h + \mathbf{H}_v^\top\mathbf{H}_v)^{-1}\\ & \times(\mathbf{u} - \mathbf{d}_2 + \mathbf{H}_h^\top(\mathbf{b}_3 + \mathbf{d}_3) + \mathbf{H}_v^\top(\mathbf{b}_4 + \mathbf{d}_4)).\end{aligned} \tag{10}$$

However, in practice this requires an inversion of a $PN \times PN$ matrix, which is intractable in most cases. But if we note that the matrices $\mathbf{H}_h$ and $\mathbf{H}_v$ are first order neighborhood gradient operators, they can be interpreted as convolution operators. Thus, these operators and their adjoint operators (whose matrices in the canonical base of $\mathbb{R}^{PN}$ are $\mathbf{H}_h^\top$ and $\mathbf{H}_v^\top$) can be computed very efficiently in the Fourier domain independently for each endmember, using their very simple convolution masks $\mathbf{h}_h$ and $\mathbf{h}_l$, and assuming periodic boundaries for the images. The masks are defined as:

$$\mathbf{h}_h = \begin{bmatrix}1 & 0 & \cdots & 0 & -1\\ 0 & 0 & \cdots & 0 & 0\\ \vdots & \vdots & & \vdots & \vdots\\ 0 & 0 & \cdots & 0 & 0\end{bmatrix} \quad \text{and} \quad \mathbf{h}_v = \mathbf{h}_h^\top. \tag{11}$$

With this in mind, using the basic properties of the Fourier transform, we give the update rule for each band p of the image $\mathcal{B}_2 \in \mathbb{R}^{m\times n\times P}$ (m and n are the vertical and horizontal dimensions of the image, such that $N = m \times n$):

$$\begin{aligned}\mathcal{B}_2^p \leftarrow \mathcal{F}^{-1}\Big(&\big(\mathcal{F}(\mathcal{U}^p - \mathcal{D}_2^p) + \mathcal{F}(\mathbf{h}_h)^* \otimes \mathcal{F}(\mathcal{B}_3^p + \mathcal{D}_3^p)\\ &+ \mathcal{F}(\mathbf{h}_v)^* \otimes \mathcal{F}(\mathcal{B}_4^p + \mathcal{D}_4^p)\big)\\ &\oslash\big(\mathbb{1}_{m\times n} + |\mathcal{F}(\mathbf{h}_h)|^2 + |\mathcal{F}(\mathbf{h}_v)|^2\big)\Big),\end{aligned} \tag{12}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ are the (discrete) 2D Fourier and inverse Fourier transforms. Each script letter corresponds to the p^{th} band of the corresponding variable represented as an $m \times n$ image. $\otimes$ and $\oslash$ are the elementwise multiplication and division, respectively. Here * is the complex conjugate and $|\cdot|$ is the complex modulus. $\mathbb{1}_{m\times n}$ is a $m \times n$ matrix of ones.

4) Optimization w.r.t. $\mathbf{b}_3$*:* Back to a matrix formulation, the optimization w.r.t. $\mathbf{B}_3$ writes:

$$\arg\min_{\mathbf{B}_3} \lambda_A||\mathbf{B}_3||_{2,1} + \frac{\rho}{2}||\mathcal{H}_h(\mathbf{B}_2) - \mathbf{B}_3 - \mathbf{D}_3||_F^2. \tag{13}$$

This problem can be rewritten in a proximal operator framework. Indeed, if for some closed proper convex function γ, we define

$$\text{prox}_\gamma(\mathbf{r}) = \arg\min_{\mathbf{s}} \gamma(\mathbf{s}) + \frac{1}{2}||\mathbf{s} - \mathbf{r}||_2^2. \tag{14}$$

as its proximal operator, then solving problem (13) is equivalent to computing:

$$\text{prox}_{(\lambda_A/\rho)||\cdot||_{2,1}}(\mathcal{H}_h(\mathbf{B}_2) - \mathbf{D}_3), \tag{15}$$

whose solution involves the *vector* or *block soft thresholding* operator, the proximal operator for the $\mathcal{L}_2$ norm:

$$\mathbf{soft}_\lambda(\mathbf{s}) = \left(1 - \frac{\lambda}{||\mathbf{s}||_2}\right)_+\mathbf{s}, \tag{16}$$

with $(\cdot)_+ = \max(\cdot, 0)$, and $\mathbf{soft}_\lambda(\mathbf{0}) = \mathbf{0}\ \forall\lambda$. This leads to the update:

$$\mathbf{B}_3 \leftarrow \mathbf{soft}_{\lambda_A/\rho}(\mathcal{H}_h(\mathbf{B}_2) - \mathbf{D}_3), \tag{17}$$

where the soft thresholding has to be understood columnwise. The horizontal gradient of $\mathbf{B}_2$ is computed in the frequency domain as in the previous section.

Note that in order to replace the spatial regularization by an anisotropic Total Variation (TV) penalization, one simply has to replace the block soft thresholding by the usual soft thresholding operator, which is the proximal operator for the $\mathcal{L}_1$ norm.

5) Optimization w.r.t. $\mathbf{b}_4$*:* Similarly, for $\mathbf{b}_4$ we have to solve:

$$\arg\min_{\mathbf{B}_4} \lambda_A||\mathbf{B}_4||_{2,1} + \frac{\rho}{2}||\mathcal{H}_v(\mathbf{B}_2) - \mathbf{B}_4 - \mathbf{D}_4||_F^2. \tag{18}$$

Using the same update rule as before:

$$\mathbf{B}_4 \leftarrow \mathbf{soft}_{\lambda_A/\rho}(\mathcal{H}_v(\mathbf{B}_2) - \mathbf{D}_4), \tag{19}$$

6) Optimization w.r.t. $\mathbf{b}_5$*:* For $\mathbf{b}_5$ the optimization problem is:

$$\arg\min_{\mathbf{B}_5} = \mathcal{I}_{\mathbb{R}_+^{P\times N}}(\mathbf{B}_5) + \frac{\rho}{2}||\mathbf{U} - \mathbf{B}_5 - \mathbf{D}_5||_F^2, \tag{20}$$

which is simply solved by

$$\mathbf{B}_5 \leftarrow (\mathbf{U} - \mathbf{D_5})_+. \tag{21}$$

7) Dual update: Finally, before going to the next iteration, the Lagrange multipliers have to be updated, giving the so-called *dual update*:

$$\mathbf{d} \leftarrow \mathbf{d} - \mathbf{\Gamma u} - \mathbf{\Lambda v}. \tag{22}$$

Then the algorithm repeats the optimization steps until a termination criterion based on primal and dual residuals (following closely [2]) is fulfilled, The barrier parameter ρ is also updated iteratively to speed up the convergence.

REFERENCES

[1] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tourneret, "Hyperspectral unmixing with spectral variability using a perturbed linear mixing model," *Signal Processing, IEEE Transactions on*, vol. 64, no. 2, pp. 525–538, 2016.

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.